



基于干扰相似度的多话题演化模型

陈叶斐, 张学军, 黄卫东
(南京邮电大学, 江苏 南京 210023)

摘要: 当前演化模型研究中, 主要是单一话题在网络中的传播, 较少考虑多话题之间的相互影响因素。在 SIR 模型的基础上提出了基于干扰相似度的多话题演化模型, 该模型中的干扰是通过话题相似度对传播概率的影响来表征的。仿真结果表明, 在临界值以内, 正负两种趋向的话题相似度分别对话题演化的进程起到加强或阻碍作用, 作用程度随着被干扰节点的度而变化, 分别表现为正向相似度下的演化一致性和负向相似度下的演化分离性。超过临界值时, 加强或阻碍作用均趋于饱和。

关键词: 话题演化; 多话题; SIR 模型; 干扰相似度

中图分类号: TP393

文献标识码: A

doi: 10.11959/j.issn.1000-0801.2017204

Multiple topics evolution model based on similarity of interference

CHEN Yefei, ZHANG Xuejun, HUANG Weidong
Nanjing University of Posts and Telecommunications, Nanjing 210023, China

Abstract: The current researches of evolution model mainly focus on the spread of the individual topics, rarely considering the influential factors between multiple topics. A new topic evolution model was proposed by considering the interference among topics based on SIR model, which characterized by the influence of the similarity of the topic on the probability of propagation. The experimental results show that within the critical value, the similarity degree of positive and negative trends enhance or hinder the process of topic evolution, and the degree of action varies with the degree of interference nodes, which is expressed as evolutionary consistency under positive similarity and the evolutionary separability under negative similarity. When the critical value is exceeded, the effect of strengthening or hindering tends to saturation.

Key words: topic evolution, multiple topics, susceptible-infective-removal model, similarity of interference

1 引言

近年来, 以互联网为代表的信息技术发展迅

猛, 互联网在全球范围内得到迅速普及。网络的发展与普及造成信息的激增, 网络个体可以更加便捷地获取信息, 并通过网络进行沟通交

收稿日期: 2017-04-01; 修回日期: 2017-06-28

基金项目: 国家自然科学基金资助项目 (No.71671093); 江苏省高校哲学社会科学基金资助项目 (No.2015SJB018); 南京邮电大学人文社会科学基金资助项目 (No.NYS214028)

Foundation Items: The National Natural Science Foundation of China(No. 71671093), Philosophy and Social Science Foundation of Education Department of Jiangsu Province(No. 2015SJB018), Humanities and Social Science Foundation of NUPT(No.NYS214028)

流。不同于报刊和广播等传统媒体,互联网表现出开放性、匿名性、交互性和跨时空性等特点,正逐渐取代传统媒体成为民众发表言论的主要渠道。

社会事件通过网络媒体的报道形成各类话题,网络用户通过参与话题的讨论和转发,与网络中的其他个体交换观点,影响各类话题的演化趋势。事件、话题和观点在网络中产生碰撞,一些讨论话题经过时间的发展和网络用户的广泛讨论、扩散,逐渐形成网络舆论。因此,话题的存在是舆论形成的前提条件,话题的演化趋势在一定程度上能够反映未来的舆论动向。

2 话题演化模型研究

近年来,研究人员针对话题演化或舆论传播构建了多种模型,并提出不同的分析方法。基于 LDA (latent dirichlet allocation) 的话题演化模型和基于复杂网络的舆论(谣言)模型是常用的两种演化模型,尽管研究对象不尽相同,但考虑到话题、舆论或谣言在网络演化中的共性,关于舆论或谣言的理论模型也可以被借鉴到话题演化的相关研究中。

LDA 模型是最早由 Blei D M^[1]提出的话题模型,在话题演化领域得到了广泛的研究。LDA 是包含文档、话题、关键词的 3 层贝叶斯概率模型,认为一篇文档中的每个词以一定概率选择了某个话题,该话题又以一定概率选择某个关键词。LDA 模型对文本中的多个话题情况进行了有效分析,但没有考虑文本信息的时间属性。在 LDA 模型的基础上,研究人员引入时间信息进行扩展,单斌等^[2]总结了基于 LDA 的扩展模型,根据时间信息的引入方式将扩展模型分为 3 类:时间信息结合型、先时间离散型和后时间离散型。

复杂网络结构模型主要有 3 类^[3]:随机网络模型、小世界模型和无标度网络模型。基于复杂

网络的舆论(谣言)演化研究多基于传染病模型,主要有 SI (susceptible-infective) 模型、SIS (susceptible-infective-susceptible) 模型、SIR (susceptible-infective-removal) 模型、SIRS (susceptible-infective-removal-susceptible) 模型、SEIR (susceptible-exposed-infective-removal) 模型和 SEIRS (susceptible-exposed-infective-removal-susceptible) 模型^[4-7],其中以 SIR 模型的应用较为广泛。目前的一种思路是基于 SIR 模型的节点状态划分,在不同的网络结构上改进 SIR 模型。Zanette D H^[8]采用 SIR 模型探究在小世界网络中的舆论传播情况,发现舆论只能传递到小于 80% 的人群。Moreno Y 等^[9]引入平均场方程,先后在均匀网络和非均匀网络上研究了舆论传播的特点。Zhou J 等^[10]基于 SIR 模型,改进了描述节点状态密度变化的演化方程,构建了一般网络舆论传播模型,得到了舆论在随机网络上最容易传播的结论。另一种思路是增加网络中的节点状态划分,基于 SIRS、SEIR、SEIRS 等模型进行演化研究。张伟^[11]在传统的 SEIR 模型的基础上,改进了演化方程和演化规则,发现改进后的模型能更好地反映舆论的演化过程。

在上述演化模型研究中,重点关注的是单一话题在网络中的传播,较少考虑多话题之间的相互影响因素^[12-15]。本文运用复杂系统的相关理论研究网络多话题演化的特征,通过网络节点、节点状态、节点间关联关系 3 个基本要素,基于 SIR 模型,考虑在多个话题的传播过程中,某一话题的演化会受到其他话题的干扰。本文通过定义不同话题间的干扰相似度,构建了基于干扰相似度的多话题演化模型,并从干扰相似度及被干扰节点的度两方面对话题演化的影响进行了仿真研究。

3 基于干扰相似度的多话题演化模型

SIR 模型是最经典的传染病模型,其描述疾



病传播的动力学微分方程组为:

$$\begin{cases} \frac{ds(t)}{dt} = -\alpha i(t)s(t) \\ \frac{di(t)}{dt} = \alpha i(t)s(t) - \beta i(t) \\ \frac{dr(t)}{dt} = \beta i(t) \end{cases} \quad (1)$$

其中, α 表示 S 态个体被感染为 I 态个体的概率, β 表示 I 态个体获得免疫转变为 S 态个体的概率。

疾病的传播强调个体的自我免疫, 个体在传播疾病的过程中获得某种免疫, 不再参与传播, 这在微分方程组中表现为感染者密度的减少和免疫者密度的增加是由于感染者以一定概率免疫 ($\beta i(t)$)。而在舆论或谣言的传播中, 传播节点通过与免疫者或传播者的接触, 受免疫者的影响或免疫者和传播者的共同影响, 以一定概率不再传播舆论或谣言, 可以认为这是一种“接触免疫”。表现在微分方程中是将 $\beta i(t)$ 用 $\beta i(t)r(t)$ 或 $\beta i(t)[i(t) + r(t)]$ 替代。传播对象的性质决定了演化特性。与疾病的传播不同, 话题的演化以网络为媒介, 以网络节点为中转, 话题在个体的交流中逐渐失去传播价值, 即话题的免疫性与免疫概率、个体状态以及个体状态对应的比例都是相关的, 所以话题停止传播可以视为接触免疫而不是个体的自我免疫。

当 SIR 模型用于话题演化的研究中时, S、I、R 这 3 种节点状态分别被重新定义为: S 态指个体了解某个话题但没有传播, 处于观望状态; I 态指个体通过对该话题的讨论交流, 进行话题传播, 处于传播状态; R 态指个体对该话题失去兴趣或被新话题吸引, 不再传播该话题或转向新话题, 处于免疫状态。

在话题演化的过程中, 观望者以一定概率转变为传播者, 传播者在遇到相同的传播者后失去传播兴趣, 或被免疫者说服转变到免疫态。对应

的演化方程为:

$$\begin{cases} \frac{ds(t)}{dt} = -P_1 s(t)i(t) \\ \frac{di(t)}{dt} = P_1 s(t)i(t) - P_R [i(t) + r(t)] \\ \frac{dr(t)}{dt} = P_R [i(t) + r(t)] \end{cases} \quad (2)$$

其中, $s(t)$ 、 $i(t)$ 、 $r(t)$ 分别表示在话题演化的过程中, t 时刻网络中观望者 S、传播者 I、免疫者 R 的节点密度; P_1 、 P_R 分别代表观望态转到传播态的传播概率、传播态转到免疫态的免疫概率, 这两个概率都是时间的函数。式 (1) 满足守恒条件 $s(t) + i(t) + r(t) = 1$, 即假设在话题演化的过程中, 参与话题传播的总节点数不变。

在多个话题的传播过程中, 某个话题的演化会受到其他话题的干扰。如果一个节点收到了某个话题的信息, 并发现周围节点都在传播相似且趋向一致的话题, 那么就会增强该节点传播该话题的动力; 相反, 如果周围节点传播的话题虽然相似, 但趋向完全相反, 该节点就会对目前接收的话题产生质疑, 就不太愿意去传播这个话题。因此, 话题间的相似度会在一定程度上影响话题传播的概率。本文的多话题干扰模型如图 1 所示, 以 2 个 S 态个体和 3 个 I 态个体为例。多个话题通过 S_1 个体与 I_2 个体间的虚线连边涌入网络中, 干扰 S_1 节点对当前话题的传播。由于网络节点在传播某个话题时, 通常会考虑周边的传播情况, 因此这种干扰在一定程度上决定了 S_1 节点能否转到传播态, 或者以多大的概率转变到传播态, 即影响了话题演化过程中的传播概率。由于干扰话题通过 S 节点与其他节点的连边涌入, 那么该 S 节点的度越大, 越容易受到外界的干扰, 因此传播概率也与节点的度相关。

定义话题演化的传播概率为:

$$P_1(t+1) = \begin{cases} 1 - [1 - P_1(t)]e^{-ave \times k}, & ave \in [0, 1] \\ P_1(t)e^{ave \times k}, & ave \in [-1, 0) \end{cases} \quad (3)$$

其中, ave 表示所有干扰话题相似度的代数平均值, k 为被干扰 S 节点的度。

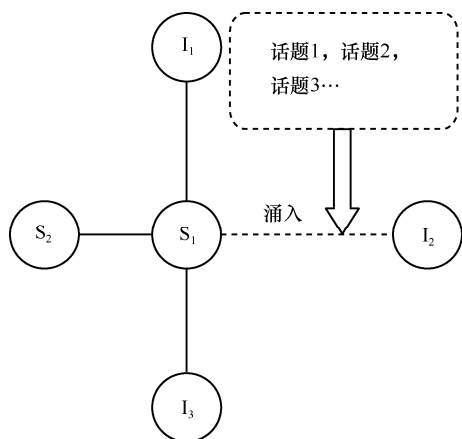


图1 多话题干扰模型

若在图1中干扰的是 I 节点, 那么免疫概率 P_R 也可以用式(3)的类似形式描述。实际仿真时只考虑对 S 节点的干扰, 将免疫概率视为常数。

为了表征不同相似度的话题干扰下的当前网络传播话题的演化效果, 定义下述参数作为衡量指标。

- 话题的生命周期 T : 指观望态、传播态和免疫态的节点密度 $s(t)$ 、 $i(t)$ 、 $r(t)$ 均达到稳定所需时间。
- 有效传播率: 用免疫者密度的稳态值 $r(T)$ 来表征。
- 话题的瞬时规模: 用传播者的瞬时密度 $i(t)$ 来表示。
- 系统的稳定性: 指话题演化达到稳定时的免疫者密度与观望者密度之比 $r(T)/s(T)$ 。
- 传播峰值: 指话题的瞬时规模取得的最大值 $i(t)_{\max}$ 。
- 峰值时刻: 指话题的瞬时规模取得最大值对应的时间 T_p 。

4 仿真结果与分析

将上述基于干扰相似度的多话题演化模型用于小世界网络, 网络模型采用 WS 小世界模型,

该模型在规则网络的基础上通过断边重连生成。小世界网络的度分布差异较小, 可近似认为是节点度分布均匀的网络。

生成节点数为 20、平均度为 8 的小世界网络。假设在话题演化的初期, 网络中只有 1 个节点处于传播态, 其余均为观望者, 不存在免疫者, 即初值为 $s(0)=0.95$, $i(0)=0.05$, $r(0)=0$ 。初始传播概率 $P_i(0)=0.3$, 免疫概率固定为 $P_i=0.1$ 。

4.1 干扰相似度对话题演化的影响

任选一节点作为被干扰的 S 节点, 度大小的平均值为 8, 若干扰话题的平均相似度 ave 分别为 0.2、-0.2, 得到对应的话题演化曲线, 并与固定概率下的话题演化进行比较, 如图 2 所示。

由图 2 可以看出, 与固定概率下的话题演化曲线相比, 正向相似 ($ave=0.2$) 的干扰话题加快了当前话题的演化进程, 而负向相似 ($ave=-0.2$) 的干扰话题对网络当前话题的演化起阻碍作用。根据图 2 中的参数进行定量计算, 得到上述定义的各项指标值, 见表 1。

从表 1 可知, 正向相似的话题干扰下, 当前话题的传播快、范围广、效率高、系统稳定好。因此, 这种干扰适合于网络话题的短期舆论造势, 话题能够快速上升为广泛的舆论。相比之下, 负向相似的话题干扰不利于话题的演化, 但如果负向相似度足够大, 就能够阻止当前话题的传播, 这为网络辟谣提供了参考。改变相似度大小, 其他仿真条件不变, 得到话题瞬时规模曲线如图 3 所示。

图 3 表明, 上述的加强或阻碍作用均对应不同的相似度临界值。由图 3(a) 可得正向相似度临界值为 0.3 左右, 由图 3(b) 可得负向相似度临界值为 -0.6 左右。一旦超过临界值, 相似话题对当前话题传播的加速或阻碍作用均趋于饱和。值得注意的是, 当负向相似度超过临界值后, 阻碍作用过大, 仅有的 1 个传播者也逐渐被说服成为免疫者, 即当前话题无法继续传播。

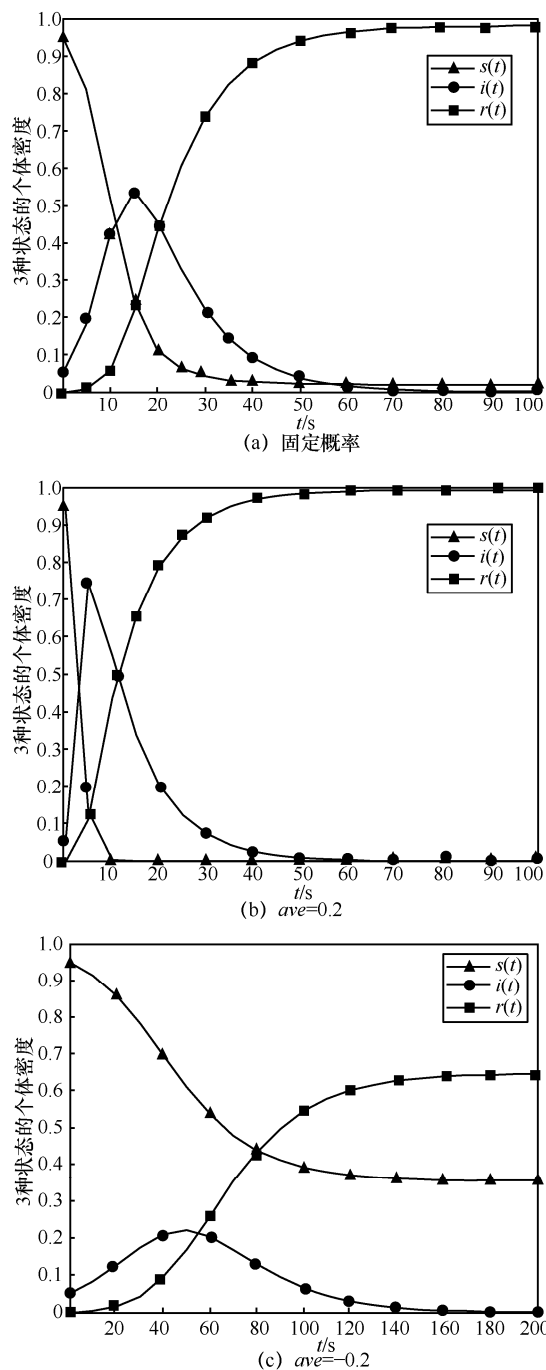


图2 不同条件下的话题演化曲线比较

表1 不同条件下的衡量指标

条件	话题生命周期	有效传播率	传播峰值	峰值时间	系统稳定性
固定概率	80	0.98	0.45	20	39.2
$ave=0.2$	60	1	0.60	10	∞
$ave=-0.2$	180	0.65	0.22	40	1.86

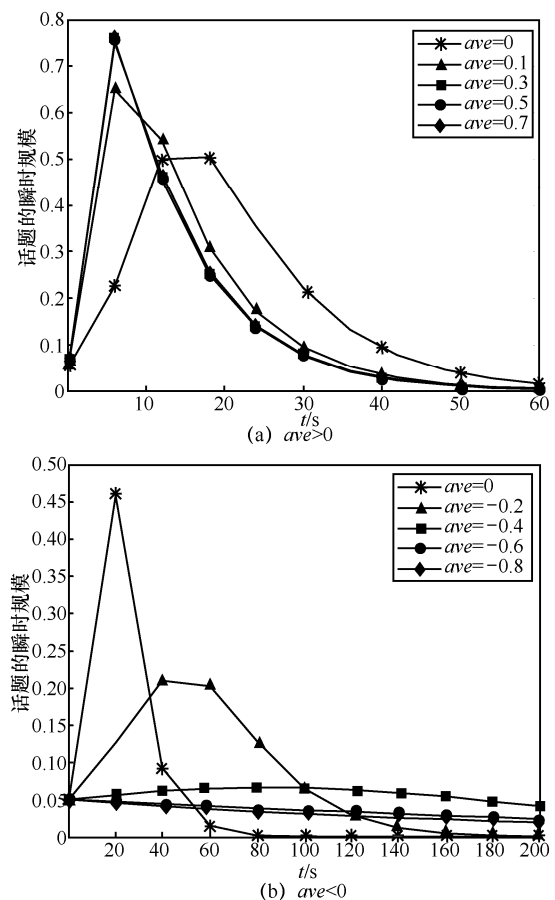


图3 话题瞬时规模随相似度的变化

4.2 被干扰节点的度对话题演化的影响

固定相似度, 改变被干扰节点的度。相似度的选取考虑到第4.1节中得到的临界值, 保证引入话题的相似度大小能够对话题演化起到明显的加强或阻碍作用。初始个体密度、初始传播概率和免疫概率不变, 得到话题瞬时规模随被干扰节点的度的变化曲线如图4和图5所示。

在图4中, 当被干扰S节点的度变化时, 话题瞬时规模曲线的重合度很高, 在同一相似度下, 话题的传播峰值相差不大, 且几乎在同一时刻取得峰值。这说明在不同正向相似的话题干扰下, 被干扰节点的度对话题瞬时规模的影响并不明显, 话题演化表现出一致性, 被干扰节点的度越大, 话题瞬时规模略有增加。话题的平均值越接近临界值0.3, 曲线的重合度越高, 此时被干扰节点的度几乎不起作用。

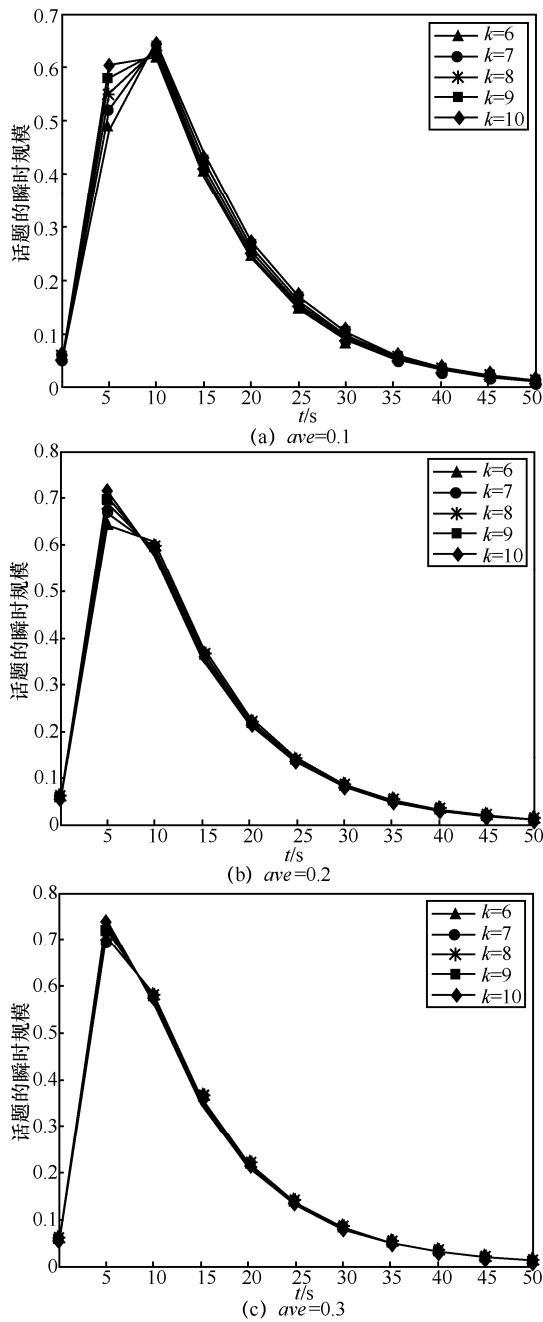


图4 话题瞬时规模随被干扰节点的度的变化 (正向)

上述实验结果表明, 在小世界网络这类节点度近似均匀的网络中, 如果采用正向相似的话题去干扰网络当前话题的演化进程, 可以忽略节点的度的差异性, 只考虑话题的相似度影响, 最终的话题演化效果相差不大。

将图5与图4相比可以发现, 当干扰话题的平均相似度为负时, 被干扰节点的度对话题演化

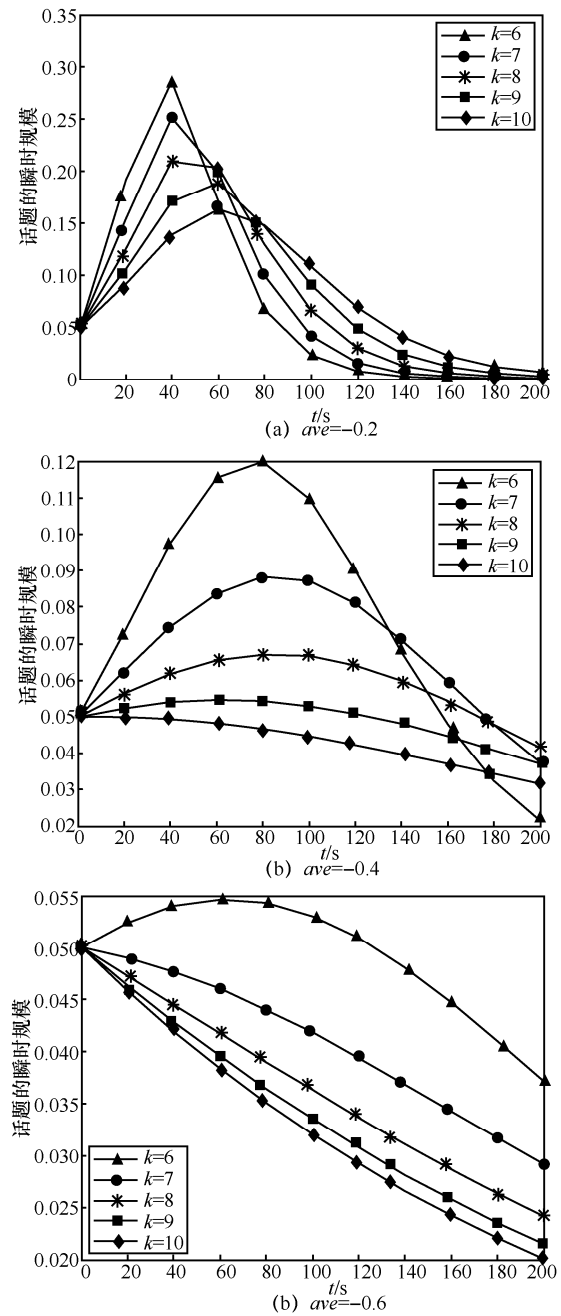


图5 话题瞬时规模随被干扰节点的度的变化 (负向)

的影响较为显著, 表现出演化分离性。在图5(a)中, 当 $ave=-0.2$ 时, 话题瞬时规模曲线的总体趋势还是相同的, 随着被干扰节点的度增大, 传播峰值减小, 峰值时间延长, 话题演化的分离程度较小。在图5(b)中, 当 $ave=-0.4$ 时, 话题演化的程度变大, 当被干扰节点的度 $k < 10$ 时, 话题演化的趋势仍维持在受到阻碍、过程减缓的情况,



与图 5 (a) 中的趋势相似; $k=10$ 时, 话题已经不再传播, 仅有的一个传播者转变成免疫者。在图 5 (c) 中, 仅在 $k=6$ 时保持图 5 (a) 中的传播趋势, 被干扰节点的度大于 6 时, 话题就停止了传播。

出现演化分离性的一个可能原因在于: 当干扰话题的相似度为负时, 节点处于停止传播前的过渡状态, 传播某个话题的动力逐渐减小。当话题的平均负向干扰相似度较小, 即使选择与外界连边较多、容易受干扰影响的节点, 只要不超过临界值, 话题演化的整体趋势不会受太大影响; 相反, 如果话题的平均负向干扰相似度较大, 即使选择的节点只比之前选择的多一条连边, 来自这一条连边的话题干扰也可能超过临界值, 迫使话题停止传播。因此, 在采用负向相似的话题去干扰当前话题的演化进程时, 被干扰节点的度的影响不可忽略。

5 结束语

在本文的模型中, 考虑了多话题间的影响因素, 探究了外来话题的相似度对话题演化过程的干扰, 为多话题模型的构建提供了一种研究方法。本文建模时只考虑了外来话题的相似度对当前话题的传播概率的影响, 对于免疫概率没有作研究, 实际仿真时将它视为常数。实际上, 如果在话题演化初期, 外来话题涌入干扰的是 I 态节点, 那么免疫概率也可以参考传播概率的定义计算式作相应修改, 可以考虑在免疫概率受话题干扰相似度下进行话题演化研究, 或者在传播概率和免疫概率二者均受到话题干扰相似度下, 研究话题演化过程。本文针对正向相似的话题干扰的实验结果, 可应用于实现话题快速而广泛的传播, 用与话题正向相似的干扰话题去影响其演化进程, 营造积极的舆论环境; 反之, 如果话题的负向相似度足够大, 可以阻止当前话题的传播, 可用于对网络中谣言或错误话题信息的控制, 进而阻止谣言的进一步扩散。

参考文献:

- [1] BLEI D M, LAFFERTY J D. Dynamic topic model[C]//The 23rd International Conference on Machine Learning, September 10-13, 2006, Pittsburgh, PA, USA. [S.l.:s.n.], 2006: 113-120.
- [2] 单斌, 李芳. 基于 LDA 话题演化研究方法综述[J]. 中文信息学报, 2010, 24(6): 43-50.
SHAN B, LI F. A survey of topic evolution based on LDA[J]. Journal of Chinese Information Processing, 2010, 24(6): 43-50.
- [3] 刘衍珩, 李飞鹏, 孙鑫, 等. 基于信息传播的社交网络拓扑模型[J]. 通信学报, 2013, 34(4): 1-9.
LIU Y H, LI F P, SUN X, et al. Social network model based on the transmission of information [J]. Journal on Communications, 2013, 34(4): 1-9.
- [4] MENG X Z, ZHAO S N, FENG T, et al. Dynamics of a novel nonlinear stochastic SIS epidemic model with double epidemic hypothesis[J]. Journal of Mathematical Analysis & Applications, 2016, 433(1): 227-242.
- [5] 朱宪莹, 刘箴, 金炜, 等. 基于特征融合的层次结构微博情感分类[J]. 电信科学, 2016, 32(7): 106-114.
ZHU X Y, LIU Z, JIN W, et al. Hierarchical micro-blog sentiment classification based on feature fusion[J]. Telecommunications Science, 2016, 32(7): 106-114.
- [6] 钟杰, 陈兴蜀, 王文贤, 等. 面向微博话题传播的重要节点测量研究[J]. 计算机应用研究, 2016, 33(8): 2290-2293.
ZHONG J, CHEN X S, WANG W X, et al. Measurement study of topology structure and crucial nodes for spread of weibo topics[J]. Application Research of Computers, 2016, 33(8): 2290-2293.
- [7] 据春华, 鲍福光, 戴俊彦. 一种融入公众情感投入分析的微博话题发现与细分方法[J]. 电信科学, 2016, 32(7): 97-105.
JU C H, BAO F G, DAI J Y. Discovery and segmentation method in micro-log topics based on public emotional engagement analysis[J]. Telecommunications Science, 2016, 32(7): 97-105.
- [8] ZANETTE D H. Dynamics of rumor propagation on small-world networks[J]. Physics, 2001, 65(1): 110-126.
- [9] MORENO Y, NEKOVEE M, VESPIGNANI A. Efficiency and reliability of epidemic data dissemination in complex networks[J]. Physical Review E Statistical Nonlinear & Soft Matter Physics, 2004, 69(50): 1-4.
- [10] ZHOU J, LIU Z H, LI B W. Influence of network structure on rumor propagation[J]. Physics Letters A, 2007, 368(6): 458-463.
- [11] 张伟. 基于复杂社会的网络舆情演化模型研究[D]. 哈尔滨: 哈尔滨工业大学, 2014.
ZHANG W. Models of the network public opinion dynamics on

complex social networks[D]. Harbin: Harbin Institute of Technology, 2014.

- [12] 孙立远, 管晓宏. 在线社会网络多话题传播竞争特性的测量[J]. 清华大学学报(自然科学版), 2015, 55(11): 1157-1162.

SUN L Y, GUAN X H. Measurements of the competitive characteristics of multi-topic propagation in online social networks[J]. Journal Tsinghua University(Sci & Technol), 2015, 55(11): 1157-1162.

- [13] 黄卫东, 林萍, 董怡, 等. 基于话题特征词的网络舆情参与者情感演化分析[J]. 情报杂志, 2015(11): 117-122, 144.

HUANG W D, LIN P, DONG Y, et al. Analysis on the feature words based evolution of netizens' sentiments in network public topics[J]. Journal of Intelligence, 2015(11): 117-122, 144.

- [14] WANG H, DENG L, XIE F, et al. A new rumor propagation model on SNS structure[C]//IEEE International Conference on Granular Computing, Aug 11-13, 2012, Hangzhou, China. New Jersey: IEEE Press, 2012: 499-503.

- [15] 王辉, 韩江洪, 邓林, 等. 基于移动社交网络的谣言传播动力学研究[J]. 物理学报, 2013, 62(11): 98-109.

WANG H, HAN J H, DENG L, et al. Dynamics of rumor spreading in mobile social networks[J]. Acta Phys Sin, 2013, 62(11): 98-109.

[作者简介]



陈叶斐(1973-), 女, 南京邮电大学助理研究员, 主要研究方向为网络舆情预警。



张学军(1969-), 男, 博士, 南京邮电大学教授, 主要研究方向为智能信息处理、复杂网络与系统和认知网络频谱感知等。



黄卫东(1968-), 男, 博士, 南京邮电大学教授, 主要研究方向为应急管理、数字化预案和网络舆情分析。